

ICR Use Cases: Louise Showe, The Wistar Institute

Use case 1:

The researcher gets the results of microarray experiment in the form of excel files containing the set of spot intensities along with the local background from all the microarrays processed in this experiment. Sample annotation with the available relevant information is provided as well in a separate file. Spot annotation with relevant gene information is also available in a separate excel file. In addition, a file containing summary of the parameters monitored during the sample preparation, hybridization, and other important steps involved as well as estimates of the original quality of the original RNA (extent of degradation, gel image, possible contamination, etc).

Upon receipt of all the information, the user performs preliminary quality control, identifies outlier and potential outlier samples, normalizes the data and proceeds to initial analysis of the data. Depending on the complexity of the experiment, the analysis is being carried over in Excel using built-in statistical functions or add-ons (such as SAM and/or PAM), using third party software, such as Statistica (www.statsoft.com), J-Express, Cluster, Treeview, or loading the data in custom-made Fortran/C/C++ analysis software. Finally, a report is prepared with the summary of the analysis (project description, including the outline of the data manipulation steps and statistical tools applied).

Use case 2.

The user is doing most (or all) of his analysis within a standard general purpose environment for data analysis (Matlab/Octave, or S-Plus/R), getting the microarray data either from a stand alone Excel/Text file with experimental data or directly from a database. The extensive use is made of built-in standard mathematical packages (routines for doing clustering, ANOVA, PCA) as well as those that were developed specifically with microarrays in mind, possibly incorporating gene annotation. The sequence of analysis steps is being automatically recorded by the software in a form suitable to be formalized as a general template for analysis of similarly designed experiments with analogous objectives, and subsequently reused.

Use case 3

The user conducted a follow-up experiment (resulting in, for example, a dataset of 30 patients and 30 controls). He would like to append it to the existing similar dataset already residing in CaArray database. Because these two experiments were carried out at different times, there is probably a systematic bias, which would have to be corrected for, before data from both datasets could be merged. In other words, the new and old datasets must be normalized with respect to one another

Use case 4

The user would like to search CaArray for microarrays done on samples that would be comparable (matching) the samples from user's new dataset. The search for matching samples would be done by matching the disease, age/sex/race of patients and controls as well as the microarray platform. Upon completion of the search, the user would like to download the retrieved data so that he could analyze them, merge them (this includes alleviating the systematic bias associated with the data being obtained from separate sources), analyze the combined dataset, and upload it to CaArray database.

Use case 5.

I have analyzed a data set and identified groups of genes whose expression patterns are highly correlated. I want to search the upstream promoter region sequence for these genes for transcription factor binding sites to identify common clusters of factor binding site in the promoter regions. I also want to see whether the correlated genes map to known pathways, whether there is information suggesting protein-protein interactions and to search the literature for papers that may have reported on co-expression of subsets of the genes.

